

Approximate evaluation of marginal association probabilities with belief propagation

Jason L. Williams ^{*†1} and Roslyn A. Lau ^{*‡2}

^{*} *Intelligence, Surveillance and Reconnaissance Division, Defence Science and Technology Organisation, Australia*

[†] *School of Electrical and Electronic Engineering, University of Adelaide, Australia*

[‡] *Research School of Computer Science, Australian National University, Australia*

¹ jason.williams@dsto.defence.gov.au, ² roslyn.lau@dsto.defence.gov.au

Abstract—Data association, the problem of reasoning over correspondence between targets and measurements, is a fundamental problem in tracking tracking. This paper presents a graphical model formulation of data association and applies an approximate inference method, belief propagation (BP), to obtain marginal association probabilities. We prove that BP is guaranteed to converge, and bound the number of iterations required for convergence. Experiments reveal a favourable comparison to prior methods in terms of accuracy and computational complexity.

Index Terms—Data association, tracking, JPDA, graphical models, belief propagation, cycles, convergence guarantees

I. INTRODUCTION

In recent years, graphical models have emerged as a powerful tool for inference and learning in large scale systems. The promise of graphical models in tracking problems was demonstrated in [1], [2], [3]. The formulation in the former focused on sensor networks, in which each sensor had a narrow field of view. Non-overlapping regions were defined and association variables were instantiated to hypothesise joint association events for all targets and measurements within a region. Graphical models were studied in a similar application in [4].

In the present study, we consider the classical data association problem, in which a single sensor surveils a large number of targets. Each target may give rise to at most one measurement, and each measurement is related to at most one target. We focus on an approximate solution of a core problem in data association: calculating marginal association probabilities such as those used in joint probabilistic data association (JPDA) [5], multi-target mixture reduction [6], [7], and related methods [8], [9].

Calculation of marginal association probabilities is closely related to computation of the permanent of a non-negative matrix [10], a key problem in the definition of the $\#P$ -complete complexity class [11]. Brute force, exact calculation of these quantities is intractable for all but the smallest problems. Until recently, practical applications of these methods have relied on simple heuristics such as those outlined in [12]. Significant improvements have been made over the past decade. The efficient hypothesis management (EHM) method [13], [14] exploits redundancy present in many problems to provide exact evaluation with reduced complexity. In many cases, however, complexity remains exponential. Monte Carlo Markov chain

data association (MCMCDA) [15] provides a randomised, fully polynomial time approximation scheme (FPTAS) for calculating the probabilities, but the computational complexity of the method limits its practical use. Subsequent to our initial publications [16], [17], a related correlation decay method [18] has been proposed, providing a deterministic FPTAS. As we show in Section V, the computational complexity of this approach remains problematic.

This paper develops a practical approximation approach to data association based on belief propagation (BP), demonstrating remarkable accuracy, and proving convergence of the algorithm, despite the presence of cycles in the graphical model formulation.

A. Contributions

In this paper, we examine an emerging method for approximating the marginal association probabilities. Motivated by the success of [19], the method was first proposed in [20], and subsequently studied in [21], [22]. It was developed independently and evaluated in the tracking context in our preliminary paper [16]. The contributions of this paper are as follows:

- Proof of convergence of the method for the most common case in tracking where the probability of detection is non-unity, and the false alarm rate is non-zero. A preliminary version of this result was published in [17]. A more general proof (effectively admitting cases with unity probability of detection and/or zero false alarm rate), developed in parallel, was announced in [23], and is available in [24].
- Analysis of the computational complexity of the method, providing guarantees on the number of iterations required as a function of the problem parameters, and interpretation of these parameters in the tracking context.
- A thorough experimental evaluation of the accuracy of the approximation in challenging tracking problems, and comparison (including computation time) to other state-of-the-art methods in the tracking literature. The comparison reveals the unique position of the proposed approximation in the accuracy vs computation time trade-off.

B. Outline of paper

In Section II, we introduce the problem of data association and the particular formulation we utilise (Section II-A), before introducing the formalism of graphical models (Section II-B). In Section III, we derive our approach, prove convergence, bound computation time, and examine practical stopping criteria. In Section IV, we compare and contrast our method to two closely related approaches. In Section V, we present a thorough experimental comparison of our approach to state-of-the-art alternatives.

II. BACKGROUND

A. Data association model

We analyse a variant of the classical model (*e.g.*, [5]) which incorporates uncertainty in target existence using the random finite set (RFS) formalism of [25]. The focus of this paper is approximate calculation of the marginal association probabilities; we describe the model for concreteness, and in order to provide intuition for how the model parameters (Section III-E) impact the number of iterations required for convergence. Details of the model and derivation of RFS filters which use the present algorithm may be found in [9].

Assume that at time t there are n_t targets with states $X_t = \{x_t^1, \dots, x_t^{n_t}\}$ and m_t measurements $Z_t = \{z_t^1, \dots, z_t^{m_t}\}$. Each target may give rise to at most one measurement (with detection probability $P_d(x_t)$), and each measurement may result from at most one target (false alarms occur according to a Poisson point process with intensity $\lambda_{fa}(z_t)$). New targets arrive at each time according to a Poisson point process with intensity $\lambda_n(x_t)$.¹ Target dynamics models (including death) are not important to the analysis we wish to perform. The measurement likelihood is $f(z_t|x_t)$. The complete set of measurements up to and including time t is denoted as $Z^t = (Z_1, \dots, Z_t)$. Under these assumptions, measurements from either false alarms or new targets will be a Poisson point process with intensity

$$\lambda_{fan}(z) = \lambda_{fa}(z) + \int f(z|x)\lambda_n(x)dx$$

The relationship between targets and measurements is described via a set of latent association variables, comprising:

- 1) For each target $i \in \{1, \dots, n_t\}$, an association variable $a_t^i \in \{0, 1, \dots, m_t\}$, the value of which is an index to the measurement with which the target is hypothesised to be associated (zero if the target is hypothesised to have not been detected)
- 2) For each measurement $j \in \{1, \dots, m_t\}$, an association variable $b_t^j \in \{0, 1, \dots, n_t\}$, the value of which is an index to the target with which the measurement is hypothesised to be associated (zero if the measurement is hypothesised to be either a false alarm or a new target)

Note that the two sets of association variables are entirely redundant: given the information from either set, the other can be reconstructed perfectly. As we will see in the following

sections, this choice of formulation results in an approximate algorithm with guaranteed convergence, and remarkable accuracy.

In essence, JPDA and related methods seek to calculate the posterior distribution

$$f(x_t^1, \dots, x_t^{n_t} | Z^t) = \sum_{a_t^1, \dots, a_t^{n_t}} p(a_t^1, \dots, a_t^{n_t} | Z^t) f(x_t^1, \dots, x_t^{n_t} | a_t^1, \dots, a_t^{n_t}, Z^t)$$

where $Z^t = (Z_1, \dots, Z_t)$. This is accomplished approximately by observing that if the prior distribution is approximated as

$$f(x_t^1, \dots, x_t^{n_t} | Z^{t-1}) = \prod_{i=1}^{n_t} f(x_t^i | Z^{t-1})$$

then, under the model stated,

$$f(x_t^1, \dots, x_t^{n_t} | a_t^1, \dots, a_t^{n_t}, Z^t) = \prod_{i=1}^{n_t} f(x_t^i | a_t^i, Z^t)$$

Subsequently, by approximating the joint distribution of association events by the product of its marginals

$$p(a_t^1, \dots, a_t^{n_t} | Z^t) \approx \prod_{i=1}^{n_t} p(a_t^i | Z^t)$$

an approximation of the full posterior can be obtained:

$$f(x_t^1, \dots, x_t^{n_t} | Z^t) \approx \prod_{i=1}^{n_t} \sum_{a_t^i} p(a_t^i | Z^t) f(x_t^i | a_t^i, Z^t)$$

Target existence can be incorporated similarly as in [9], [26], or using the related methods [27]. The hypothesis-conditioned updated distribution $f(x_t^i | a_t^i, Z^t)$ may be calculated using simple state estimation methods such as a Kalman filter [28], unscented Kalman filter [29] or particle filter [30] depending on the model in use. The challenging problem is computation of the marginal probability $p(a_t^i | Z^t)$; this computation is #P complete. Under the assumptions stated, the joint distribution can be expressed as

$$p(a_t^1, \dots, a_t^{n_t}, b_t^1, \dots, b_t^{m_t} | Z^t) \propto \prod_{i=1}^{n_t} \left\{ \psi_i(a_t^i) \prod_{j=1}^{m_t} \psi_{i,j}(a_t^i, b_t^j) \right\} \quad (1)$$

where the factors

$$\psi_{i,j}(a_t^i, b_t^j) = \begin{cases} 0, & a_t^i = j, b_t^j \neq i \text{ or } b_t^j = i, a_t^i \neq j \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

collectively ensure that the redundant sets of association variables $(a_t^1, \dots, a_t^{n_t})$ and $(b_t^1, \dots, b_t^{m_t})$ are consistent, *i.e.*, that any event in which the collections are inconsistent has zero probability. An example of an inconsistent event is one in which the target association variable a_t^i indicates that target i is associated with measurement j , but the measurement association variable b_t^j does not indicate that measurement j is associated with target i ; this implicitly excludes cases in which the same measurement is associated with two targets, or the same target is associated with two measurements. The

¹For simplicity, we assume that new targets are guaranteed to be detected; the method in [9] shows how this assumption can be relaxed.

factors ψ_i encode the problem data, with $\psi_i(a_t^i = 0) = 1$ and for $a_t^i = j > 0$,

$$\psi_i(j) = \frac{q_t^i \int P_d(x_t^i) f(z_t^j | x_t^i) f(x_t^i | Z^{t-1}) dx_t^i}{\lambda_{\text{fan}}(z_t^j) [1 - q_t^i \int P_d(x_t^i) f(x_t^i | Z^{t-1}) dx_t^i]} \quad (3)$$

where q_t^i is the probability of existence of the i -th track prior to incorporation of the new measurement.² We reinforce that the details of this model are not the topic of this paper; for the present context, the model is a minor variation of previous works such as [5], [7], [8]. The topic of interest is how we might tractably estimate the marginal association probabilities

$$p(a_t^i | Z^t) = \sum_{a_t^{i'}, i' \neq i; b_t^j \forall j} p(a_t^1, \dots, a_t^{n_t}, b_t^1, \dots, b_t^{m_t} | Z^t) \quad (4)$$

$$p(b_t^j | Z^t) = \sum_{a_t^i \forall i; b_t^{j'}, j' \neq j} p(a_t^1, \dots, a_t^{n_t}, b_t^1, \dots, b_t^{m_t} | Z^t) \quad (5)$$

B. Graphical models

Graphical models [31], [32], [33] aim to represent and manipulate the joint probability distributions of many variables efficiently by exploiting factorisation. The Kalman filter [34] and the hidden Markov model (HMM) [35] are two examples of algorithms that exploit sparsity of a particular kind (*i.e.*, a Markov chain) to efficiently conduct inference on systems involving many random variables. Inference methods based on the graphical model framework generalise these algorithms to a wider variety of state spaces and dependency structures.

Graphical model methods have been developed for undirected graphical models (Markov random fields), directed graphical models (Bayes nets) and factor graphs. In this work we consider a subclass of pairwise undirected models, involving nodes (*i.e.*, random variables) $n \in \mathcal{N}$, and edges (*i.e.*, dependencies) $e \in \mathcal{E} \subset \mathcal{N} \times \mathcal{N}$, and where the joint distribution can be written as:³

$$p(x_{\mathcal{N}}) \propto \prod_{n \in \mathcal{N}} \psi_n(x_n) \prod_{(i,j) \in \mathcal{E}} \psi_{i,j}(x_i, x_j)$$

It should be immediately clear that (1) is in this form. As another example, a Markov chain involving variables (x_1, \dots, x_T) may be formulated by setting $\psi_1(x_1) = p(x_1)$ for the initial prior, $\psi_t(x_t) = 1$ for $t > 1$, and edges $\psi_{t-1,t}(x_{t-1}, x_t) = p(x_t | x_{t-1})$, $t \in \{2, \dots, T\}$ representing the Markov transition kernels, although other formulations are possible.

Optimal inference can be conducted on tree-structured graphs using belief propagation (BP). BP proceeds by passing messages between neighbouring nodes. We denote by $\mu_{i \rightarrow j}(x_j)$ the message sent from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$ where $(i, j) \in \mathcal{E}$. The iterative update equations are then:

$$\mu_{i \rightarrow j}(x_j) \propto \sum_{x_i} \psi_{i,j}(x_i, x_j) \psi_i(x_i) \prod_{(j', i) \in \mathcal{E}, j' \neq j} \mu_{j' \rightarrow i}(x_i) \quad (6)$$

²We do not consider update equations for this quantity, hence we do not require notation for the post-update probability of existence.

³In the general setting, the joint distribution is a product of maximal cliques [32, p9]. Since the graph is undirected, we assume that \mathcal{E} is symmetric, *i.e.*, if $(i, j) \in \mathcal{E}$ then $(j, i) \in \mathcal{E}$. We need only incorporate one of these two factors in the distribution.

For obvious reasons, the algorithm is also known as *sum-product*. If the summations are replaced with maximisation operations, then we arrive at max-product BP, which generalises the well-known Viterbi algorithm [36], providing the MAP joint state of all variables in the graph. At convergence of sum-product BP, the marginal distribution at a node n can be calculated as:

$$p(x_n) \propto \psi_n(x_n) \prod_{(n,i) \in \mathcal{E}} \mu_{i \rightarrow n}(x_n) \quad (7)$$

In the case of a Markov chain, if all nodes are jointly Gaussian, then BP is equivalent to a Kalman smoother. Similarly, if all nodes are discrete, then BP is equivalent to inference on an HMM using the forwards-backwards algorithm. BP unifies these algorithms, and extends them from chains to trees.

Inference in cyclic graphs (graphs that have cycles, *i.e.*, that are not tree-structured) is far more challenging. Conceptually, one can always convert an arbitrary cyclic graph to a tree by merging nodes (*e.g.*, so-called *junction tree* representations) [31], [33], but in practical problems, the dimensionality of the agglomerated variables may be prohibitive. BP may be applied to cyclic graphs; practically, this simply involves repeated application of (6) until convergence occurs (*i.e.*, until the maximum error between subsequent messages is less than a pre-set threshold). Unfortunately, this is neither guaranteed to converge to the right answer, nor to converge at all. Nevertheless, and perhaps surprisingly, it has exhibited excellent empirical performance in many practical problems [37]. For example, the popular iterative turbo decoding algorithm has been shown to be an instance of BP applied to a cyclic graph [38].

The current understanding of BP in cyclic graphs stems largely from [39]. It has been shown (*e.g.*, [32, Theorem 3.4]) that one can recover exact marginal probabilities from an optimisation of a convex function known as the Gibbs free energy, one term of which is the joint entropy of the distribution. While this optimisation is intractable, it points to a family of *variational inference* methods that approximate the objective function and the feasible set to enable calculation of marginal probability estimates without ever manipulating the full joint distribution [40], [32]. It has been shown [39] that BP (when it converges) solves one such approximation, in which the objective is approximated by the Bethe free energy (replacing the joint entropy with a series of differences of pairwise edge entropies and node entropies), and the feasible set (*i.e.*, the set of all valid probability distributions) is approximated as the distributions for which the pairwise joint distributions along edges are consistent with node marginals [32, 4.1.1]. The BP message iterates can be viewed as a general iterative method for solving a series of fixed point equations derived from the optimality conditions of the Bethe free energy variational problem [32, 4.1.3].

The two approximations made by BP each lead to difficulties in certain circumstances. Firstly, whereas the Gibbs free energy is convex (since entropy is concave), the Bethe free energy is, in general, neither concave nor convex. The practical behaviour is that errors are often either very small or very large, and the general intuition is that failure of convexity

leads to large errors; this has led to new families of algorithms such as tree re-weighted sum product [41], which replace the Bethe free energy with a convex surrogate. Secondly, the approximation of the feasible set may admit points which do not correspond to any valid probability distribution, hence the solution may be infeasible. Methods seeking to tighten the feasible set include [42].

In the case in which the graph is a tree, Bethe free energy is equivalent to Gibbs free energy (and hence convex), and the feasible set is exact. Not surprisingly, then, BP is optimal on trees. In recent years, a number of other cases have emerged in which there are guarantees on aspects of the algorithm; this paper studies one of these cases. In [19], it was shown that max-product LBP can be used to optimally solve 2D assignment problems in time comparable to the well-known auction algorithm. This paper concerns the calculation of marginal probabilities on the same model, and utilises the same graph formulation. In [24], it was shown that, in this model, Bethe free energy can be parameterised by the doubly stochastic matrix of marginal target-measurement probabilities, and is convex with respect to this parameterisation. Furthermore, the Birkhoff-von Neumann theorem states that any doubly-stochastic matrix is a convex combination of permutations [24] (e.g., different association configurations), hence infeasibility is also ruled out.⁴ In the following section, we state the model, and derive results relating to convergence and complexity of BP in this model.

III. BELIEF PROPAGATION DATA ASSOCIATION

In this section, we consider the use of BP to approximate the association probabilities in (1). We commence in Section III-A by explicitly stating the BP update equations, and then (in Section III-B) obtain an equivalent form that reduces computation complexity from $O(n_t^3 m_t^2 + n_t^2 m_t^3)$ per iteration to $O(n_t m_t)$. The reduction is effectively a sum-product version of the simplified algorithm in [19], and an equivalent form was provided in [20]. In Section III-C we use the simplified update equations to prove convergence of the algorithm, and then, in Section III-D, bound the number of iterations required for convergence. In Section III-E we calculate the expected value of the quantity involved in the bound on the number of iterations required, providing intuition into the practical behaviour of the method. In Section III-F we provide a criterion for terminating the BP computation with a guarantee of the deviation from the converged solution.

A. Formulation

The model we study is that of (1). As illustrated in Fig. 1, this is a bipartite model in which all target association variables $(a_t^i)_{i \in \{1, \dots, n_t\}}$ are connected to all measurement association variables $(b_t^j)_{j \in \{1, \dots, m_t\}}$. In this case, BP may be implemented via two half-iterations, alternating between the two sets of messages, $(\mu_{a_t^i \rightarrow b_t^j})$ and $(\mu_{b_t^j \rightarrow a_t^i})$; abbreviating notation, we refer to these as $(\mu_{i \rightarrow j})$ and $(\mu_{j \rightarrow i})$ respectively. The message update equations are:

⁴i.e., there is guaranteed to be a distribution over joint associations that yields the marginal probabilities that BP obtains.

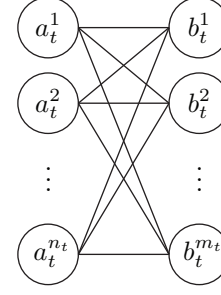


Fig. 1. Graphical model formulation employed for data association. The value of the random variable a_t^i is the index of the measurement with which target i is hypothesised to be associated, while the value of the random variable b_t^j is the index of the target with which measurement j is hypothesised to be associated.

$$\mu_{i \rightarrow j}(b_t^j) = \sum_{a_t^i} \psi_i(a_t^i) \psi_{i,j}(a_t^i, b_t^j) \prod_{j' \neq j} \mu_{j' \rightarrow i}(a_t^i) \quad (8)$$

$$= \begin{cases} \psi_i(j) \prod_{j' \neq j} \mu_{j' \rightarrow i}(j), & b_t^j = i \\ \sum_{a_t^i \neq j} \psi_i(a_t^i) \prod_{j' \neq j} \mu_{j' \rightarrow i}(a_t^i), & b_t^j \neq i \end{cases} \quad (9)$$

$$\mu_{j \rightarrow i}(a_t^i) = \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j) \prod_{i' \neq i} \mu_{i' \rightarrow j}(b_t^j) \quad (10)$$

$$= \begin{cases} \prod_{i' \neq i} \mu_{i' \rightarrow j}(i), & a_t^i = j \\ \sum_{b_t^j \neq i} \prod_{i' \neq i} \mu_{i' \rightarrow j}(b_t^j), & a_t^i \neq j \end{cases} \quad (11)$$

Naïvely, the complexity per iteration of this procedure is $O(n_t^3 m_t^2 + n_t^2 m_t^3)$, since each iteration involves sending a message from each of the n_t target association variables to each of the m_t measurement association variables (and vice versa), and each message involves $(m_t + 1)$ (respectively $(n_t + 1)$) values, each of which requires $O(n_t^2)$ (respectively $O(m_t^2)$) calculations.

B. Simplified algorithm

As shown in [19], [20], [17], these computations can be dramatically simplified by observing that in (9) (and (11)), while the message consists of $(n_t + 1)$ (respectively $(m_t + 1)$) values there are only two distinct values (i.e., $b_t^j = i$ and $b_t^j \neq i$ in (9), and $a_t^i = j$ and $a_t^i \neq j$ in (11)). Further, since we are free to renormalise messages, we may divide by one of these two values ($\mu_{i \rightarrow j}(b_t^j \neq i)$ and $\mu_{j \rightarrow i}(a_t^i \neq j)$ respectively) to obtain a scalar representation of the message. The message update equations in terms of these scalars become:

$$\mu_{i \rightarrow j} = \frac{\psi_i(j)}{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i}} \quad (12)$$

$$\mu_{j \rightarrow i} = \frac{1}{1 + \sum_{i' \neq i, i' > 0} \mu_{i' \rightarrow j}} \quad (13)$$

where we have exploited the choice that $\psi_i(a_t^i = 0) = 1$ to reinforce that the denominator is greater than zero. Upon convergence, the approximate marginal association probabilities

can be obtained by

$$\hat{p}(a_t^i = j | Z^t) = \frac{\psi_i(j) \mu_{j \rightarrow i}}{\sum_{j'} \psi_i(j') \mu_{j' \rightarrow i}} \quad (14)$$

$$\hat{p}(b_t^j = i | Z^t) = \frac{\mu_{i \rightarrow j}}{\sum_{i'} \mu_{i' \rightarrow j}} \quad (15)$$

where $\mu_{0 \rightarrow i} \triangleq 1$ and $\mu_{0 \rightarrow j} \triangleq 1$. An $O(n_t m_t)$ (per iteration) Matlab implementation of these update equations was given in [17].

C. Proof of convergence

Most previous work on convergence of BP (such as [43]) requires the factors of the graph to satisfy conditions on dynamic range (e.g., $\max_{a,b} \psi_{i,j}(a,b) / \min_{a,b} \psi_{i,j}(a,b)$). These methods do not apply to the present case since the factor $\psi_{i,j}(a,b)$ has infinite dynamic range. Nevertheless, we show here that it is possible to establish that the simplified expressions in Eqs. (12) and (13) are contractions, thus guaranteeing convergence of the method. The property we exploit in the proof is the same as the correlation decay property utilised in the recent work [18].

A function $f : \mathcal{X} \rightarrow \mathcal{X}$ operating in a metric space \mathcal{X} with distance metric $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ is a *contraction* if there is a contraction factor $\alpha < 1$ such that $d(f(x), f(y)) \leq \alpha d(x, y) \forall x, y \in \mathcal{X}$ [44, 9.22]. If f is a contraction on a complete metric space, then any sequence resulting from repeated application of f will converge to the same fixed point [44, 9.23]. Following [43], we use the distance metric

$$d(\mu, \tilde{\mu}) = \max_{i,j} \left| \log \frac{\mu_{i \rightarrow j}}{\tilde{\mu}_{i \rightarrow j}} \right| \quad (16)$$

where we use the convention $\frac{0}{0} = 1$.⁵ To show that this satisfies the triangle inequality, note that

$$\begin{aligned} d(\mu, \tilde{\mu}) &= \max_{i,j} \left| \log \frac{\mu_{i \rightarrow j}}{\nu_{i \rightarrow j}} + \log \frac{\nu_{i \rightarrow j}}{\tilde{\mu}_{i \rightarrow j}} \right| \\ &\leq \max_{i,j} \left| \log \frac{\mu_{i \rightarrow j}}{\nu_{i \rightarrow j}} \right| + \max_{i,j} \left| \log \frac{\nu_{i \rightarrow j}}{\tilde{\mu}_{i \rightarrow j}} \right| \\ &= d(\mu, \nu) + d(\nu, \tilde{\mu}) \end{aligned}$$

Other properties of a distance metric are trivially satisfied.

Let $\mu = \mathbf{f}(\nu)$ be the update defined in (12) in vector form with $\nu = (\mu_{j \rightarrow i})_{i \in \{1, \dots, n_t\}, j \in \{1, \dots, m_t\}}$, $\mu = (\mu_{i \rightarrow j})_{i \in \{1, \dots, n_t\}, j \in \{1, \dots, m_t\}}$, and $\nu = \mathbf{g}(\mu)$ be the update defined in (13). We now prove a preliminary result relating to the form of the contraction factor that we use.

Lemma 1. For $L > 1$ and $c > 0$, the function

$$\alpha(L, c) = \frac{\log \left(\frac{1+cL}{1+c} \right)}{\log L} \quad (17)$$

is strictly less than one, and is monotonically increasing in L .

⁵In our problem setup, elements will either be bounded below by a strictly positive constant, or zero at all points in the sequence (when $\psi_i(j) = 0$), hence we will never measure the distance between a zero element and a non-zero element (which would be infinite).

Proof: First note that

$$\frac{1+cL}{1+c} = L \frac{\frac{1}{L} + c}{1+c} < L$$

This gives the first result. For the second result, it suffices to show that $\frac{d}{dL} \log \alpha(L, c) \geq 0$, i.e., that

$$\left(\frac{1+cL}{1+c} \right) \log \left(\frac{1+cL}{1+c} \right) \leq \frac{c}{1+c} L \log L$$

This result is an immediate consequence of convexity of $x \log x$, since $\frac{1+cL}{1+c} = \frac{1}{1+c} + \frac{c}{1+c} L$. ■

The following two lemmas establish that the BP updates in Eqs. (12) and (13) are contractions.

Lemma 2. For all $(\nu, \tilde{\nu})$ with $d(\nu, \tilde{\nu}) \leq \log \bar{L}$, the message update $\mathbf{f}(\cdot)$ is a contraction with respect to the distance metric $d(\cdot, \cdot)$ with factor $\alpha(\bar{L}, W_*)$, where

$$W_i = \sum_{a_i^i > 0} \psi_i(a_i^i), \quad W_* = \max_i W_i$$

Proof: Assume that, $\forall (i, j)$, $0 < \mu_{j \rightarrow i} \leq 1$ and $0 < \tilde{\mu}_{j \rightarrow i} \leq 1$; this is not restrictive as it is guaranteed to be satisfied for any $(\nu, \tilde{\nu})$ resulting from (13). Let

$$L \triangleq \exp d(\nu, \tilde{\nu}) \leq \bar{L} < \infty$$

Then $\mu_{j \rightarrow i} \leq L \tilde{\mu}_{j \rightarrow i}$ and $\tilde{\mu}_{j \rightarrow i} \leq L \mu_{j \rightarrow i}$. If $\psi_i(j) = 0$ then $f_{i,j}(\nu) = f_{i,j}(\tilde{\nu}) = 0$, and $f_{i,j}(\nu)/f_{i,j}(\tilde{\nu}) \triangleq 1$. Otherwise, consider the quotient

$$\begin{aligned} \frac{f_{i,j}(\nu)}{f_{i,j}(\tilde{\nu})} &= \frac{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \tilde{\mu}_{j' \rightarrow i}}{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i}} \\ &\leq \frac{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') L \mu_{j' \rightarrow i}}{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i}} \\ &= \frac{1+cL}{1+c} \leq \frac{1+W_*L}{1+W_*} \\ &= L^{\alpha(L, W_*)} \leq L^{\alpha(\bar{L}, W_*)} \end{aligned}$$

where $c = \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i} \leq W_i \leq W_*$, and the final step uses the second result of Lemma 1. Following similar steps,

$$\frac{f_{i,j}(\tilde{\nu})}{f_{i,j}(\nu)} \leq \frac{1+W_*L}{1+W_*} \leq L^{\alpha(\bar{L}, W_*)}$$

After combining these two cases, taking logs and observing that they apply for all (i, j) , we are done. ■

We now show the same result for the equation in the alternative step, (13).

Lemma 3. For all $(\mu, \tilde{\mu})$ with $d(\mu, \tilde{\mu}) \leq \log \bar{L}$, the message update $\mathbf{g}(\cdot)$ is a contraction with respect to the distance metric $d(\cdot, \cdot)$ with factor $\alpha(\bar{L}, W^*)$, where

$$W^j = \sum_{i > 0} \psi_i(j), \quad W^* = \max_j W^j$$

Proof: Assume that, $\forall (i, j)$, $0 \leq \mu_{i \rightarrow j} \leq \psi_i(j)$ and $0 \leq \tilde{\mu}_{i \rightarrow j} \leq \psi_i(j)$, and that $\mu_{i \rightarrow j} = 0$ or $\tilde{\mu}_{i \rightarrow j} = 0$ if and only if $\psi_i(j) = 0$ (in which case $\mu_{i \rightarrow j} / \tilde{\mu}_{i \rightarrow j} \triangleq 1$). Again, this is not restrictive as it is guaranteed to be satisfied for any $(\mu, \tilde{\mu})$

resulting from (12). Consequently, $\sum_i \mu_{i \rightarrow j} \leq W^j \leq W^*$ (and similarly for $\tilde{\mu}$). Let

$$L \triangleq \exp d(\mu, \tilde{\mu}) \leq \bar{L} < \infty$$

Then, following the same steps as in the previous proof,

$$\begin{aligned} \frac{g^{i,j}(\mu)}{g^{i,j}(\tilde{\mu})} &= \frac{1 + \sum_{i' \neq i} \tilde{\mu}_{i' \rightarrow j}}{1 + \sum_{i' \neq i} \mu_{i' \rightarrow j}} \\ &\leq \frac{1 + W^* L}{1 + W^*} \leq L^{\alpha(\bar{L}, W^*)} \\ \frac{g^{i,j}(\tilde{\mu})}{g^{i,j}(\mu)} &\leq \frac{1 + W^* L}{1 + W^*} \leq L^{\alpha(\bar{L}, W^*)} \end{aligned}$$

Since this is true for all (i, j) , we take logs, combine cases, and obtain the desired result. ■

Theorem 1. *The loopy BP fixed point iteration $\mathbf{f}(\cdot)$, $\mathbf{g}(\cdot)$ converges to the same stationary point regardless of the initialisation.*

We omit the proof of the theorem since it is a straightforward application of contraction mapping results; see [44, Sect 9.23], and note that the message iterates are contained in a compact subset of the space (i.e., with elements either being zero for all k , or being bounded below and above by strictly positive finite constants) hence establishing completeness.

D. Bound on complexity

Having proven convergence of the algorithm, we now switch to analysing the computational complexity involved in attaining convergence. Since we are ultimately interested in the marginal probability estimates, we start with a lemma that relates the deviation in the BP messages from their final values to the deviation in the marginal probability estimates (we use the term deviation to describe the difference between the current iterate and the converged solution). For convenience we define the shorthand $p^i(j) \triangleq \hat{p}(a_t^i = j | Z^t)$.

Lemma 4. *If $\left| \log \frac{\mu_{j \rightarrow i}}{\tilde{\mu}_{j \rightarrow i}} \right| \leq \epsilon \forall (i, j)$ and p and \tilde{p} are calculated from μ and $\tilde{\mu}$ via (14), then $|p^i(j) - \tilde{p}^i(j)| \leq \delta(\epsilon) \forall (i, j)$ where $\delta(\epsilon) = \exp(2\epsilon) - 1 \approx 2\epsilon$ (for small ϵ).*

Proof: We seek to bound

$$\begin{aligned} &\left| \frac{\mu_{j \rightarrow i} \psi_i(j)}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} - \frac{\tilde{\mu}_{j \rightarrow i} \psi_i(j)}{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')} \right| \\ &= \left| \frac{\mu_{j \rightarrow i} \psi_i(j)}{\tilde{\mu}_{j \rightarrow i} \psi_i(j)} \cdot \frac{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} - 1 \right| \cdot \frac{\tilde{\mu}_{j \rightarrow i} \psi_i(j)}{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')} \\ &\leq \left| \frac{\mu_{j \rightarrow i}}{\tilde{\mu}_{j \rightarrow i}} \cdot \frac{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} - 1 \right| \end{aligned} \quad (18)$$

Trivially, $\exp(-\epsilon) \leq \frac{\mu_{j \rightarrow i}}{\tilde{\mu}_{j \rightarrow i}} \leq \exp(\epsilon)$. The second term can be bounded similarly since

$$\frac{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} \leq \frac{\sum_{j'} \exp(\epsilon) \mu_{j' \rightarrow i} \psi_i(j')}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} = \exp(\epsilon)$$

Substituting these bounds into (18) we find

$$\exp(-2\epsilon) - 1 \leq \frac{\mu_{j \rightarrow i}}{\tilde{\mu}_{j \rightarrow i}} \cdot \frac{\sum_{j'} \tilde{\mu}_{j' \rightarrow i} \psi_i(j')}{\sum_{j'} \mu_{j' \rightarrow i} \psi_i(j')} - 1 \leq \exp(2\epsilon) - 1$$

The desired result is then obtained by observing that $0 \leq 1 - \exp(-2\epsilon) \leq \exp(2\epsilon) - 1$. Finally, note that $\exp(2\epsilon) - 1 = 2\epsilon + O(\epsilon^2)$ for small ϵ , so that the error is well-approximated as 2ϵ . ■

Using the previous lemma, we now show that the number of iterations required for the deviation to be less than a desired level is bounded through a simple closed-form expression. In what follows, we denote by $\mu_{i \rightarrow j, k}$ and $\mu_{j \rightarrow i, k}$ the k -th iterate of the messages from (12) and (13), and by $\mu_{i \rightarrow j, *}$ and $\mu_{j \rightarrow i, *}$ the messages upon convergence.

Theorem 2. *Starting from $\mu_{j \rightarrow i, 0} = 1$ and given W_* and W^* , deviation between the marginal probabilities and the approximations obtained at convergence is guaranteed to be no more than ϵ if the number of iterations k satisfies*

$$k - 1 \geq \frac{\log[\exp(2\epsilon) - 1] - \log \log(1 + W_*)}{\log \alpha(1 + W_*, W_*) + \log \alpha(1 + W_*, W^*)}$$

Looser bounds that only consider the contractions corresponding to one half-message also apply:

$$k - 1 \geq \frac{\log[\exp(2\epsilon) - 1] - \log \log(1 + W_*)}{\log \alpha(1 + W_*, W_*)} \quad (19)$$

$$k \geq \frac{\log[\exp(2\epsilon) - 1] - \log \log(1 + W^*)}{\log \alpha(1 + W^*, W^*)} \quad (20)$$

Proof: First, we bound the distance between the messages at initialisation. The point which we choose to bound is the message $\mu_{i \rightarrow j, 0}$ which results from a single application of (12) from the initialisation $\mu_{j \rightarrow i, 0} = 1$. We examine the ratio

$$\frac{\mu_{i \rightarrow j, 0}}{\mu_{i \rightarrow j, *}} = \frac{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i, *}}{1 + \sum_{j' \neq j, j' > 0} \psi_i(j') \mu_{j' \rightarrow i, 0}}$$

It is clear that any $\mu_{j \rightarrow i}$ resulting from (13) will satisfy $0 < \mu_{j \rightarrow i} \leq 1$. Consequently,

$$\frac{1}{1 + W_*} \leq \frac{1}{1 + W_i} \leq \frac{\mu_{i \rightarrow j, 0}}{\mu_{i \rightarrow j, *}} \leq 1$$

Thus, after half an iteration,

$$\left| \log \frac{\mu_{i \rightarrow j, 0}}{\mu_{i \rightarrow j, *}} \right| \leq \log(1 + W_*)$$

Thus we set $\bar{L} = 1 + W_*$, to obtain a contraction factor $\alpha = \alpha(1 + W_*, W_*)\alpha(1 + W_*, W^*)$ (combining the two half-iterations). After one more half-iteration (neglecting the contraction that this introduces for convenience), we obtain:

$$\left| \log \frac{\mu_{j \rightarrow i, 1}}{\mu_{j \rightarrow i, *}} \right| \leq \log(1 + W_*)$$

After $(k - 1)$ subsequent iterations, the deviation will be bounded by

$$\left| \log \frac{\mu_{j \rightarrow i, k}}{\mu_{j \rightarrow i, *}} \right| \leq \alpha^{k-1} \log(1 + W_*)$$

In order to obtain the desired bound on the marginal probability deviation, we seek k such that

$$\alpha^{k-1} \log(1 + W_*) \leq \exp(2\epsilon) - 1$$

$$(k - 1) \log \alpha \leq \log[\exp(2\epsilon) - 1] - \log \log(1 + W_*)$$

$$k - 1 \geq \frac{\log[\exp(2\epsilon) - 1] - \log \log(1 + W_*)}{\log \alpha}$$

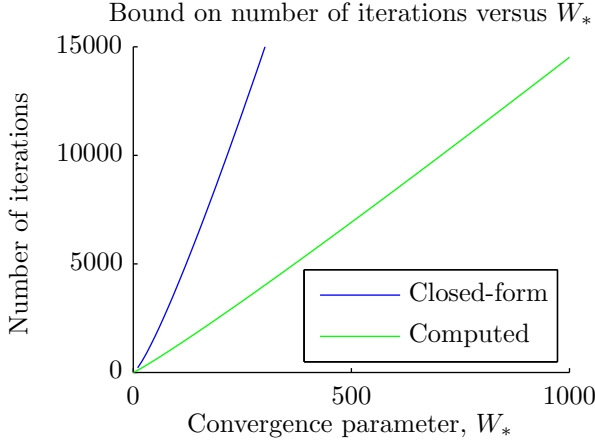


Fig. 2. Bounds on the number of iterations required for convergence (*i.e.*, to guarantee that the resulting marginal estimates are within 10^{-3} of their final values). The blue line shows the closed-form bound of (19), while the green line shows the tighter bound computed using Corollary 1.

The first loosened bound is obtained by observing that $\alpha \leq \alpha(1 + W_*, W_*)$, *i.e.*, only incorporating the contraction factor for one of the two half-iterations. The second is obtained by observing that

$$\frac{1}{1 + W_*} \leq \mu_{j \rightarrow i, *} \leq 1$$

and following similar subsequent steps. ■

The analysis in Theorem 2 can be tightened via a bound that is computed numerically. As iterations progress, we are guaranteed to be closer to the optimal solution. Correspondingly, the contraction factor reduces, hastening convergence. The closed-form analysis does not exploit this fact, rather assuming that the contraction factor remains at its original, worst-case value. The tighter, computable bound is described in Corollary 1.

Corollary 1. Assume that we commence from $\mu_{j \rightarrow i, 0} = 1$, so that

$$\left| \log \frac{\mu_{i \rightarrow j, 1}}{\mu_{i \rightarrow j, *}} \right| \leq \log(1 + W_*) \triangleq L_1$$

Then for $k > 1$

$$\left| \log \frac{\mu_{i \rightarrow j, k}}{\mu_{i \rightarrow j, *}} \right| \leq L_k$$

where

$$L_k = \alpha(\exp L_{k-1}, W_*) L_{k-1}$$

Fig. 2 shows the number of iterations k required to guarantee deviation $\delta = 10^{-3}$ in the marginal estimates for various W_* ranging in $[0, 1000]$. The green line shows the computable bound from Corollary 1, while the blue line shows the closed-form bound from Theorem 2 (*i.e.*, the first loosened bound, which depends only on W_*). The computable bound is well-approximated as being log-log-linear, while the closed-form bound is well-approximated as being log-linear.

E. Interpretation of complexity

The previous section showed that the number of iterations required depends on the parameters $W_* = \max_i W_i$ and/or $W^* = \max_j W^j$. In this section, we evaluate the expected value $E_{Z|X}[W_i]$ when the true target positions are $X = \{x^1, \dots, x^n\}$ in order to provide intuition on how this parameter relates to the problem parameters. Modifying notation to explicitly incorporate the measurement set Z ,

$$W_i = \sum_{z \in Z} \psi_i(z)$$

Simplifying the model to a uniform probability of detection, false alarm density and new target density, the weights $\psi_i(z)$ become

$$\psi_i(z) = \frac{q_t^i P_d}{\lambda_{\text{fan}}(1 - q_t^i P_d)} \int f(z|x_t^i) f(x_t^i|Z^{t-1}) dx_t^i$$

The sum in W_i is over all measurements; accordingly its expected value can be calculated using the first moment of the measurement distribution [45], [25] to obtain

$$E_{Z|X}[W_i] = \int \psi_i(z) \lambda(z|X) dz \quad (21)$$

$$= \frac{q_t^i P_d}{\lambda_{\text{fan}}(1 - q_t^i P_d)} \int \lambda(z|X) f^i(z|Z^{t-1}) dz \quad (22)$$

where $f^i(z|Z^{t-1})$ is the distribution of the measurement for the i -th track, and $\lambda(z|X)$ is the first moment of the measurement distribution given the true multi-target state X :

$$f^i(z|Z^{t-1}) = \int f(z|x_t^i) f(x_t^i|Z^{t-1}) dx_t^i \quad (23)$$

$$\lambda(z|X) = \lambda_{\text{fan}} + \sum_{i=1}^n P_d f(z|x^i) \quad (24)$$

Accordingly, we interpret (22) as being the expected measurement intensity in the vicinity of the predicted measurement distribution for the track i . This provides useful intuition on the dependence on various problem parameters:

- W_i decreases as the false alarm intensity increases. The reduction is (approximately) linear when the false alarm intensity is a small contributor to the overall measurement intensity. When the false alarm intensity is the dominant contributor to the intensity, little reduction will occur.
- W_i decreases as the probability of existence and detection decreases. This occurs both due to the leading factor, and due to the reduction in the overall measurement intensity. As the probability of existence and detection approaches unity, the quantity increases rapidly, due to the term $q_t^i P_d / (1 - q_t^i P_d)$ in the leading factor.
- W_i increases as targets become closely spaced, as multiple targets contribute to the measurement intensity in the region for the track of interest. Highly accurate measurements and dynamic models will reduce W_i if the additional accuracy permits targets to be distinguished, *i.e.*, it reduces the contribution of other targets. If the spacing of the targets is such that the additional accuracy does not separate the targets, an increase in accuracy will increase W_i , increasing convergence time.

In summary, if additional SNR permits disambiguation of measurement-target association then it will aid convergence; in other cases, lower SNR cases will yield better convergence. Empirically, we will see in Section V that the marginal probability estimates also become closer to their exact values in lower SNR conditions. A similar analysis will show that the quantities W^* and W^j depend on the intensity of tracks in the vicinity of measurements, although the quantities involved are not as easily interpreted.

F. Stopping criterion

The emphasis of the previous sections has been on bounding the number of iterations required to obtain an error less than a particular threshold. Practically, this will lead to an algorithm that performs a pre-determined number of iterations. The more common approach in LBP is to compare messages to previous iterates and terminate when convergence is detected. The following theorem, which is a well-known application of contraction mapping results, provides a stopping criterion with a guaranteed bound on error.

Theorem 3. *If $d(\mu_k, \mu_{k-1}) \leq \epsilon$ then*

$$d(\mu_k, \mu_*) \leq \frac{\alpha(\epsilon, W_*)\alpha(\epsilon, W^*)}{1 - \alpha(\epsilon, W_*)\alpha(\epsilon, W^*)}\epsilon$$

Proof: Let μ_k denote the vector form of the messages at iteration k and $\alpha = \alpha(\epsilon, W_*)\alpha(\epsilon, W^*)$. Then $\forall k' \geq k$, $d(\mu_{k'}, \mu_{k'-1}) \leq \alpha^{k'-k}\epsilon$. By repeated application of the triangle inequality,

$$d(\mu_k, \mu_*) \leq \sum_{k'=k+1}^{\infty} d(\mu_{k'}, \mu_{k'-1}) \leq \sum_{l=1}^{\infty} \alpha^l \epsilon = \frac{\alpha}{1 - \alpha} \epsilon$$

■

G. Algorithm

The BP algorithm is summarised in Fig. 3. The algorithm incorporates the results of Lemma 4 and Theorem 3 in order to test convergence. Convergence checks are performed every N iterations in order to avoid the computational overhead involved. In practice, we suggest values in the range $N \in \{5, \dots, 20\}$.

IV. RELATIONSHIP TO OTHER METHODS

A. Junction tree

The *junction tree* algorithm [46], [33] is the standard method for conducting exact inference in a cyclic graph. The algorithm provides a systematic procedure for merging⁶ variables into *hyper-nodes* in order to convert the cyclic graph into a tree, and then executes BP on that tree to conduct exact inference. As previously described, the complexity of the algorithm can be problematic, as the computation increases exponentially in

⁶loosely speaking; more correctly, the method forms a tree of hyper-nodes that satisfies the running intersection property, thus ensuring that enforcing consistency along edges is sufficient. For further details, see [33].

Require: Number of tracks n_t , number of measurements m_t , single-target association weights $\psi_i(j) \forall i \in \{1, \dots, n_t\}, j \in \{1, \dots, m_t\}$ (assumes $\psi_i(0) = 1$), convergence criterion δ , number of iterations between convergence checks N

Ensure: Marginal probability estimates $p_{ij} \forall i \in \{1, \dots, n_t\}, j \in \{0, \dots, m_t\}$

$\eta := \frac{1}{2} \log(1 + \delta)$ {For convergence criterion}

$W_* := \max_i \sum_{j>0} \psi_i(j)$ {For convergence criterion}

$\mu_{j \rightarrow i} := 1 \forall i, j > 0$

repeat

{Perform N iterations without checking convergence}

for $k := 1$ **to** N **do**

for $i := 1$ **to** n_t **do** {Calculate L-R messages}

$s := 1 + \sum_{j>0} \psi_i(j) \mu_{j \rightarrow i}$

$\mu_{i \rightarrow j} := \psi_i(j) / [s - \psi_i(j) \mu_{j \rightarrow i}] \forall j$

end for

if $k = N$ **then** {For convergence check}

$\tilde{\mu}_{j \rightarrow i} := \mu_{j \rightarrow i} \forall i, j > 0$

end if

for $j := 1$ **to** m_t **do** {Calculate R-L messages}

$s := 1 + \sum_{i>0} \mu_{i \rightarrow j}$

$\mu_{j \rightarrow i} := 1 / [s - \mu_{i \rightarrow j}] \forall i$

end for

end for

{Check for convergence}

$d := \max_{i,j} \left| \log \frac{\mu_{j \rightarrow i}}{\tilde{\mu}_{j \rightarrow i}} \right|$

$\alpha := \left(\log \frac{1+W_*d}{1+W_*} \right) / (\log d)$

until $\frac{\alpha d}{(1-\alpha)} < \eta$

{Calculate marginal estimates}

for $i := 1$ **to** n_t **do**

$s := 1 + \sum_{j>0} \psi_i(j) \mu_{j \rightarrow i}$

$p_{i,j} := \psi_i(j) \mu_{j \rightarrow i} / s \forall j > 0$

$p_{i,0} := 1/s$

end for

Fig. 3. Algorithm for computing approximate marginal probabilities using BP.

the number of nodes that need to be merged together in order to yield a tree-structured graph.

Junction trees can be applied to data association using the graph described in Fig. 1, but this would be quite inefficient, since the redundant use of target association variables and measurement association variables is of no benefit in the exact case, and it increases the number of variables over which inference must be conducted. Instead, we use the simpler, yet equivalent representation of (1):

$$p(a_t^1, \dots, a_t^{n_t} | Z^t) \propto \left[\prod_{i=1}^{n_t} \psi_i(a_t^i) \right] \cdot \left[\prod_{(i,i') \in \mathcal{E}} \psi_c(a_t^i, a_t^{i'}) \right] \quad (25)$$



Fig. 4. Example of graphical formulation used as an input to the junction tree algorithm. In this case, there is no measurement with non-zero weight for both target 1 and target 4, hence there is no edge between the corresponding nodes.

where

$$\psi_c(a_t^i, a_t^{i'}) = \begin{cases} 0, & a_t^i = a_t^{i'} > 0 \\ 1, & \text{otherwise} \end{cases} \quad (26)$$

$$\mathcal{E} = \{(i, i') | \exists i \in \{1, \dots, n_t\}, i' \in \{1, \dots, n_t\}, j \in \{1, \dots, m_t\} \text{ s.t. } i \neq i', \psi_i(j)\psi_{i'}(j) > 0\} \quad (27)$$

We assume that the node potentials $\psi_i(j)$ have been thresholded such that they are zero for infeasible associations. This in turn creates the sparsity in \mathcal{E} which is exploited by the junction tree algorithm. While BP could also be applied directly to the cyclic graph in this formulation (as proposed in [47], [33, Box 12.D]), it was shown in [16] that it performs poorly (in both accuracy and convergence) in comparison to the bipartite formulation in (1). An example of this graph is shown in Fig. 4.

The efficient hypothesis management method (EHM-2) [13], [14] exploits a similar tree-based inference, gaining additional efficiency by effectively reducing the alphabet within the junction tree hyper-nodes (*i.e.*, eliminating elements of the hyper-node alphabet which violate mutual exclusion constraints).

The experiments in Section V compare the accuracy and computation time of the method proposed in this paper to the junction tree algorithm with various thresholds applied to $\psi_i(j)$, yielding various levels of sparsity in the resulting graph. The experiments were conducted using the implementation of junction tree contained in the library for discrete approximate inference [48].

B. Correlation decay

The recently proposed method of [18] utilises the correlation decay property of statistical physics to obtain a deterministic FPTAS for the marginal association probabilities. The algorithm involves a recursion of the form

$$\Phi(\mathcal{T}, \mathcal{M}, i, t) = \frac{1}{1 + \sum_{j \in \mathcal{M}} \psi_i(j) \Phi(\mathcal{T} - \{i\}, \mathcal{M}, j, t-1)} \quad (28)$$

$$\Phi(\mathcal{T}, \mathcal{M}, j, t) = \frac{1}{1 + \sum_{i \in \mathcal{T}} \psi_i(j) \Phi(\mathcal{T}, \mathcal{M} - \{j\}, i, t-1)} \quad (29)$$

where \mathcal{T} (respectively, \mathcal{M}) is the set of tracks (respectively, measurements) remaining in the recursion, and t is the maximum number of recursion steps to perform. Close inspection reveals that these equations are almost identical to (12) and (13); the difference is that the recursion in (28) and (29) never revisits nodes in its recursion; consequently it is exact, but has exponential complexity.

The theoretical analysis in [18] shows that if the number of nodes in the graph grows, but the maximum connectivity and maximum single-target association weight remains constant, t may be chosen such that the error is bounded yet complexity is polynomial in the number of tracks and measurements. The proof of this theorem, which is based on the earlier work [49], exploits similar properties to the convergence proof in Section III-C.

The experiments in Section V compare the accuracy and computation time of the correlation decay algorithm to the method proposed in this paper.

V. EXPERIMENTS

We consider a series of single time-step problems involving targets on a regularly spaced grid. Although the experiment involves only a single time step, the track covariances are preinitialised by simulating 30 time steps of the standard constant velocity model:

$$\begin{aligned} P_{t|t-1} &= F P_{t-1|t-1} F^T + Q \\ P_{t|t} &= \begin{cases} P_{t|t-1}, & \text{target not detected} \\ P_{t|t-1} - K H P_{t|t-1}, & \text{target detected} \end{cases} \\ K &= P_{t|t-1} H^T (H P_{t|t-1} H^T + R)^{-1} \end{aligned}$$

with

$$F = I_{2 \times 2} \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad Q = q I_{2 \times 2} \otimes \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix},$$

$T = 1$, $q = 0.01$, $H = I_{2 \times 2} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix}$, and $R = r I_{2 \times 2}$, commencing with $P_{0|0}$ equal to zero. The track estimates were initialised to the true target positions, corrupted with additive Gaussian noise with covariance distributed according to the prior covariance for the target, generated through independent simulation of each target. Gating was performed with a threshold such that the probability of excluding the target-derived measurement was 10^{-4} . The area populated with false alarms was sufficiently large to cover these gates.

Three groups of experiments were conducted:

- 1) Six targets (arranged in a regular 2×3 grid), varying target spacing between 0 and 10 units, considering cases with $P_d \in \{0.3, 0.6, 0.9\}$, $\lambda_{fa} \in \{0.01, 0.0316, 0.1\}$, and $r \in \{0.1, 1, 10\}$
- 2) Targets in a $n \times 3$ regular grid (with $n \in \{2, \dots, 30\}$) with target spacing set to 3 units, $P_d = 0.6$, $\lambda_{fa} = 0.01$ and $r = 1$
- 3) Targets in a $n \times n$ regular grid (with $n \in \{2, \dots, 10\}$) with target spacing set to 3 units, $P_d = 0.6$, $\lambda_{fa} = 0.01$ and $r = 1$

A number of example experiments are shown in Fig. 5.

For each parameter (spacing, P_d , etc), 1000 single time step Monte Carlo trials were performed. The total number of Monte

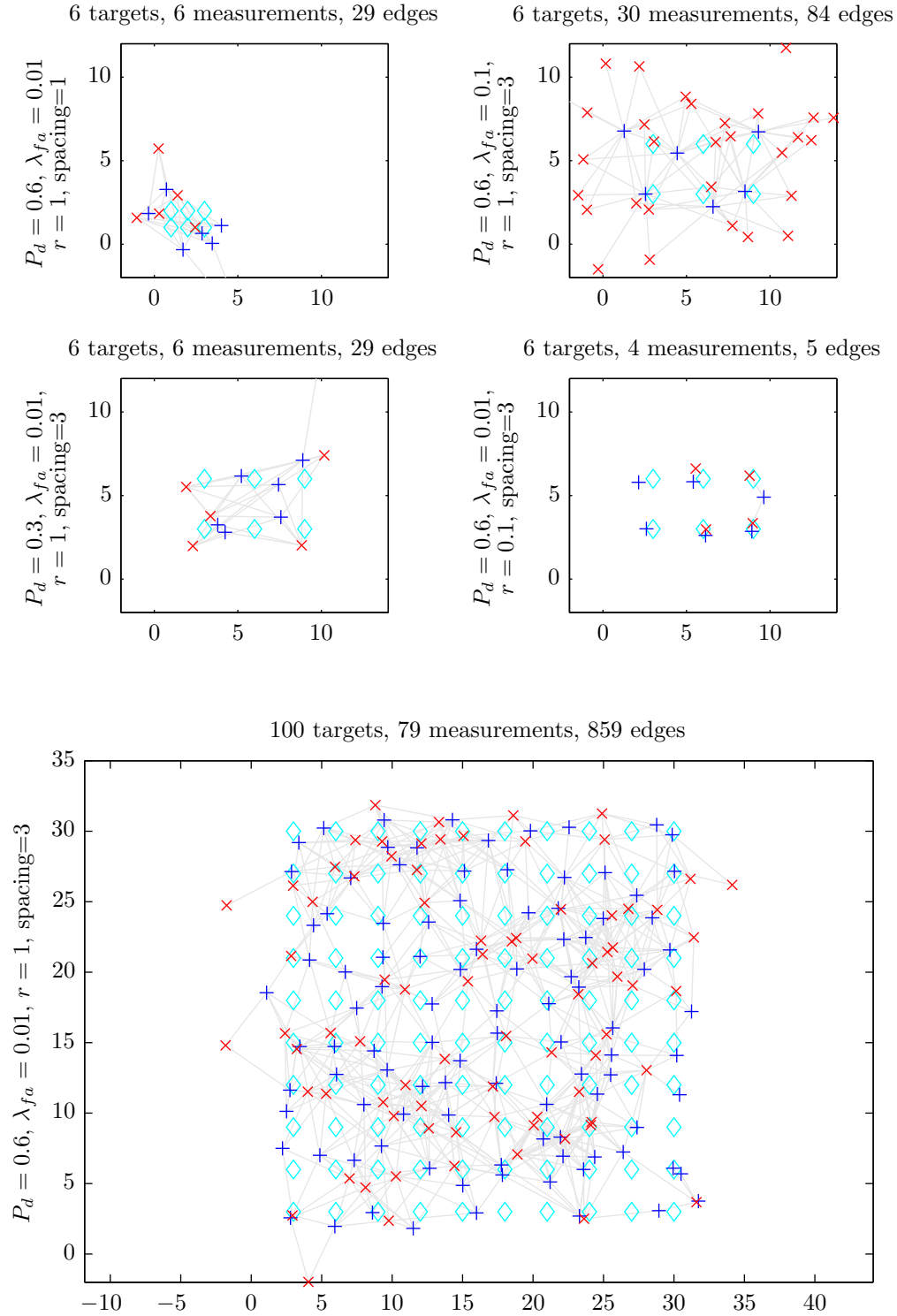


Fig. 5. Examples of single Monte Carlo trials. True target positions are shown as \diamond , measurements (false alarms and true measurements) are shown as \times , and prior track estimates are shown as $+$. Feint grey lines show the edges in the graph, connecting track estimates and measurements which form feasible associations.

Carlo trials performed for each algorithm was 395,000. The experiments were performed using a dual processor Intel Xeon E5-2670 server with Matlab Parallel Computing Toolbox, utilising 12 worker threads.

To evaluate the various algorithms compared, we calculated the average maximum error in the marginal probabilities

calculated by each algorithm, and the average computation time. The comparison is performed on the basis of the accuracy of the marginal estimates in order to exclude down-stream effects such as coalescence, mixture reduction, *etc.* While these effects are significant, they are separate issues to the approximation of the marginal probabilities, and any conclusion

reached in an experiment including these effects is specific to the full gamut of approximations made in the tracking system. We define the average maximum error as being the average value of the largest element of the difference between the marginal distribution of a target association variable estimated by the algorithm under test, and the corresponding reference value, thus averaging is performed over both Monte Carlo trials and over targets in those trials. In the first experiment, the reference value is the exact marginal distribution, calculated using the junction tree algorithm with a threshold of 10^{-4} (which is applied to the weights utilised in all algorithms). The size of the problems in the second and third experiments prevented exact computation, thus the reference values were calculated using MCMCDA with 10^7 samples. A second instance of this algorithm was included as a test, providing an estimate of the expected error in the reference values.

A. Comparison algorithms

The algorithms compared in the experiments were:

- The BP algorithm detailed in Fig. 3. A vectorised, Matlab-based implementation was utilised.
- The junction tree algorithm described in Section IV-A, with weights thresholded to 10^{-3} , 10^{-2} and 10^{-1} , inducing varying degrees of sparsity, and thus exploring the accuracy versus computation time trade-off. The algorithm was implemented using the general-purpose libDAI system [48].
- The MCMCDA algorithm of [15], using 10^5 , 10^6 and 10^7 MCMC steps. The algorithm was initialised with the MAP association (calculated via an auction) to avoid the need for burn-in. The implementation was written in Matlab and compiled into C++ using the Matlab Coder.
- The correlation decay algorithm described in Section IV-B, using $t \in \{3, 5, 7\}$. The implementation was written in C++.
- The linear multitarget integrated existence PDA (LMIPDA) algorithm of [50]. The track existence probabilities were set to unity for comparison to non-IPDA methods. A vectorised, Matlab-based implementation was utilised.
- The approximate Bakhtiar-Alavi-Amoozegar (BAA) algorithm from [51], which was shown to be the approximation of choice in [12]. A vectorised, Matlab-based implementation was utilised.

The preliminary version of this paper [16] included results for PDA (*i.e.*, ignoring the presence of adjacent targets), and BP applied directly to the cyclic graph in the alternative formulation in Fig. 4 (as discussed in Section IV-A). These were excluded from this comparison as the errors they committed were considerably worse than the algorithms we are comparing.

B. Results and discussion

The results of the comparison are shown in Fig. 6. The rows of plots alternate between average marginal error, and average computation time (per simulation). Plots with shaded

backgrounds correspond to the second and third experiments (*i.e.*, varying numbers of targets), while the remaining plots correspond to the six target experiments. The x -axis in the six-target experiments shows the spacing of the true positions of the targets in the regular grid, varying between 0 and 10 units. The first two rows of plots show results varying the false alarm rate with $\lambda_{fa} \in \{0.01, 0.0316, 0.1\}$, with $P_d = 0.6$ and $r = 1$. The first case $\lambda_{fa} = 0.01$ is the baseline, to which the following cases compare. The second two rows show the effect of lowering or raising the P_d , *i.e.*, $P_d \in \{0.3, 0.9\}$. The final two rows show the effect of raising or lowering the measurement noise, *i.e.*, $r \in \{0.1, 10\}$. The top two plots in the shaded box show the second experiment (with $3 \times n$ targets varying $n \in \{2, \dots, 30\}$, where the x -axis shows the total number of targets), and the third experiment (with $n \times n$ targets varying $n \in \{2, 10\}$, again with the x -axis showing the total number of targets). In each of these latter cases, the target spacing is fixed to 3 units, and baseline parameters apply.

We make the following observations:

- BP (blue) exhibits excellent performance, with average errors of 0.015 or less in most cases, except for the high P_d case ($P_d = 0.9$) and the low measurement noise case ($r = 0.1$); in these two cases, the errors are in the 0.02 – 0.04 range. Note that these two cases correspond to high SNR. The large scale problems ($3 \times n$ and $n \times n$) show little or no indication that the accuracy of BP deteriorates as the number of targets increases.
- BP (blue) significantly outperforms LMIPDA (magenta) and BAA (cyan), exhibiting errors reduced by a factor of ten in most cases. The computational complexity of BP is only a few times that of LMIPDA and BAA, averaging fractions of a millisecond in all of the six-target problems, and a little over a millisecond even in the 90 and 100 target problems.
- LMIPDA (magenta) and BAA (cyan) appear perform similarly to each other across the board. BAA performs significantly better in high false alarm and high measurement noise cases, while LMIPDA performs better in low measurement noise cases. The computation time is comparable.
- Junction tree (black) exhibits excellent performance with smaller thresholds, but at a high computational cost. With the threshold set to 0.1 (black dash-dotted), the error is globally worse than BP, and often much worse. The computational complexity in this case is 1-2 orders of magnitude higher than BP, and the 9×9 and 10×10 experiments were unable to be completed due to the high memory requirements of the algorithm. With the threshold set to 0.01 or 0.001, the accuracy of junction tree is generally much better than BP, but the computation time is 2-4 orders of magnitude higher than BP. Most experiments in the $3 \times n$ and $n \times n$ cases were unable to be completed due to excessive memory requirements.
- The performance of MCMCDA (green) with 10^5 iterations (dot-dashed) is generally comparable to BP in the six-target cases, but it performs progressively worse in cases with larger number of targets. In the larger cases,

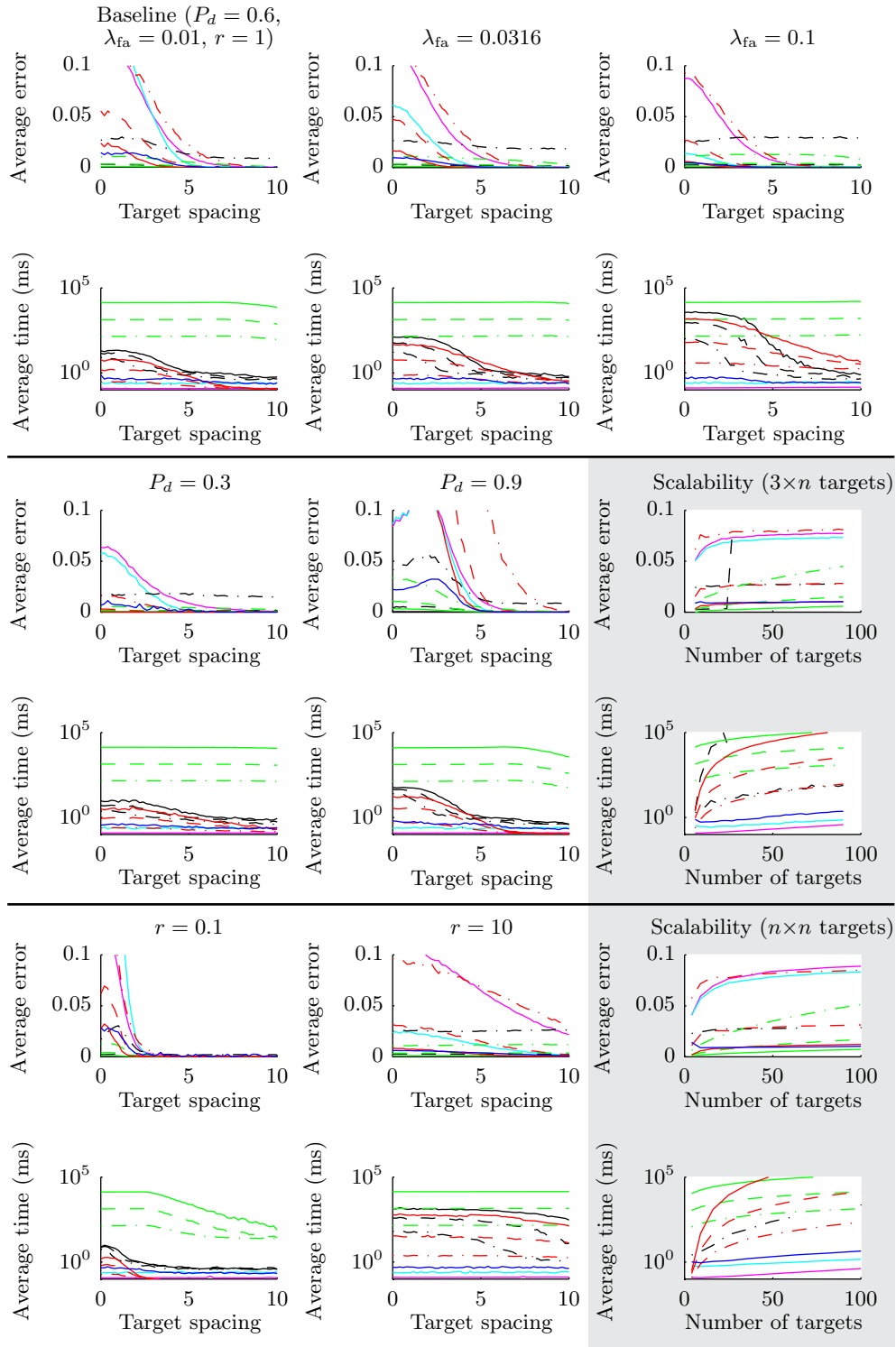


Fig. 6. Results of experiments. First, third and fifth rows of plots show the average worst-case error in the marginal probabilities, averaged over targets and Monte Carlo trials. Second, fourth and sixth rows show the average computation time. Unshaded plots show results of the first experiment (involving six targets on a regular grid with spacing varied on the x -axis), while shaded plots show the second and third experiments (with $3 \times n$ and $n \times n$ targets and respectively). Colours show algorithms as BP, LMIPDA, BAA, MCMCDA (with dot-dashed, dashed and solid representing 10^5 , 10^6 and 10^7 iterations), correlation decay (with dot-dashed, dashed and solid representing 3, 5 and 7 iterations) and junction tree (with dot-dashed, dashed and solid representing thresholds of 10^{-1} , 10^{-2} and 10^{-3}).

BP performs similar to or better than MCMCDA with 10^6 iterations. The complexity of MCMCDA is 2-4 orders of magnitude higher than BP depending on the number of samples used.

- BP generally outperforms the correlation decay method (red) with $t = 3$ (dot-dashed) and $t = 5$ (dashed). With $t = 7$, BP generally performs better for tighter target spacings, and correlation decay performs better

with larger target spacings. The correlation decay method performs particularly poorly in the case with $P_d = 0.9$. The computational complexity of the correlation decay method varies greatly through the experiments. In some cases (e.g., $r = 0.1$) it is slightly faster than BP. In many other cases its complexity is 2-3 orders of magnitude slower, and in the large scale problems, it is 4-5 orders of magnitude slower.

VI. CONCLUSION

This paper has introduced a new approximate method for solving data association problems using BP on a particular graphical model formulation. While BP is generally not guaranteed to converge on cyclic graphs, we have proven convergence on the formulation studied, and bounded the computation required to attain convergence. While the method is approximate, the experiments in Section V reveal an highly favourable comparison with state-of-the-art methods in the accuracy versus computation time trade-off.

ACKNOWLEDGEMENTS

The authors thank Prof Alan Willsky and Prof Müjdat Çetin for many helpful discussions during development of the formulation in Section III.

REFERENCES

- [1] L. Chen, M. J. Wainwright, M. Çetin, and A. S. Willsky, "Multitarget-multisensor data association using the tree-reweighted max-product algorithm," in *Proc SPIE Signal Processing, Sensor Fusion, and Target Recognition*, vol. 5096, August 2003, pp. 127–138.
- [2] L. Chen, M. Çetin, and A. S. Willsky, "Distributed data association for multi-target tracking in sensor networks," in *Proc. 8th International Conference on Information Fusion*, July 2005.
- [3] L. Chen, M. J. Wainwright, M. Çetin, and A. S. Willsky, "Data association based on optimization in graphical models with application to sensor networks," *Mathematical and Computer Modelling*, vol. 43, no. 9–10, pp. 1114–1135, 2006.
- [4] A. Gning and L. Mihaylova, "Dynamic clustering and belief propagation for distributed inference in random sensor networks with deficient links," in *Proc. 12th International Conference on Information Fusion*, July 2009, pp. 656–663.
- [5] K.-C. Chang and Y. Bar-Shalom, "Joint probabilistic data association for multitarget tracking with possibly unresolved measurements and maneuvers," *IEEE Trans. Autom. Control*, vol. 29, no. 7, pp. 585–594, July 1984.
- [6] L. Y. Pao, "Multisensor multitarget mixture reduction algorithms for tracking," *Journal of Guidance, Control, and Dynamics*, vol. 17, no. 6, pp. 1205–1211, 1994.
- [7] B. Ristic and S. Arulampalam, "Multitarget mixture reduction algorithm with incorporated target existence recursions," *Proc. SPIE Signal and Data Processing of Small Targets*, vol. 4048, pp. 357–368, July 2000.
- [8] P. Horridge and S. Maskell, "Searching for, initiating and tracking multiple targets using existence probabilities," in *Proc. 12th International Conference on Information Fusion*, Seattle, WA, July 2009, pp. 611–617.
- [9] J. L. Williams, "Graphical model approximations of random finite set filters," arXiv, e-print arXiv:1105.3298, 2011. [Online]. Available: <http://arxiv.org/abs/1105.3298>
- [10] J. Collins and J. Uhlmann, "Efficient gating in data association with multivariate Gaussian distributed states," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 909–916, Jul. 1992.
- [11] L. G. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science*, vol. 8, no. 2, pp. 189–201, 1979.
- [12] K. Romeo, D. F. Crouse, Y. Bar-Shalom, and P. Willett, "The JPDAF in practical systems: approximations," in *Signal and Data Processing of Small Targets*, O. E. Drummond, Ed. Orlando, FL: SPIE, 2010, vol. 7698.
- [13] S. Maskell, M. Briers, and R. Wright, "Fast mutual exclusion," in *Proc SPIE Signal and Data Processing of Small Targets*, vol. 5428, 2004, pp. 526–536.
- [14] P. Horridge and S. Maskell, "Real-time tracking of hundreds of targets with efficient exact JPDAF implementation," in *Proc. 9th International Conference on Information Fusion*, July 2006, pp. 1–8.
- [15] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for multi-target tracking," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 481–497, Mar. 2009.
- [16] J. L. Williams and R. A. Lau, "Data association by loopy belief propagation," in *Proc. 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.
- [17] —, "Convergence of loopy belief propagation for data association," in *Proc. 6th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Brisbane, Australia, December 2010, pp. 175–180.
- [18] S. Oh, "A distributed deterministic approximation algorithm for data association," in *Proc. International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, June 2011.
- [19] M. Bayati, D. Shah, and M. Sharma, "Max-product for maximum weight matching: Convergence, correctness, and LP duality," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1241–1251, March 2008.
- [20] B. Huang and T. Jebara, "Approximating the permanent with belief propagation," arXiv, e-print arXiv:0908.1769v1, August 2009. [Online]. Available: <http://arxiv.org/abs/0908.1769>
- [21] M. Chertkov, L. Kroc, and M. Vergassola, "Belief propagation and beyond for particle tracking," arXiv, e-print arXiv:0806.1199v1, June 2008. [Online]. Available: <http://arxiv.org/abs/0806.1199>
- [22] M. Chertkov, L. Kroc, M. Vergassola, and L. Zdeborová, "Inference in particle tracking experiments by passing messages between images," arXiv, e-print arXiv:0909.4256v2, May 2010. [Online]. Available: <http://arxiv.org/abs/0909.4256>
- [23] P. O. Vontobel, "The Bethe permanent of a non-negative matrix," in *Proc. 48th Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, IL, September/October 2010, pp. 341–346.
- [24] —, "The Bethe permanent of a non-negative matrix," arXiv, e-print arXiv:1107.4196v1, July 2011. [Online]. Available: <http://arxiv.org/abs/1107.4196>
- [25] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.
- [26] J. L. Williams, "Experiments with graphical model implementations of multiple target multiple Bernoulli filters," in *Proc. 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Adelaide, Australia, December 2011, pp. 532–537.
- [27] D. Mušicki and R. J. Evans, "Joint integrated probabilistic data association: JIPDA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1093–1099, July 2004.
- [28] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. Academic Press, Inc, 1979, vol. 1.
- [29] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
- [30] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [31] S. L. Lauritzen, *Graphical models*. Oxford, UK: Clarendon Press, 1996, vol. 17.
- [32] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [33] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.
- [34] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [35] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.
- [36] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, April 1967.
- [37] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. 15th Conference on Uncertainty in Artificial Intelligence*, 1999, pp. 467–476.
- [38] R. McEliece, D. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 140–152, Feb 1998.

- [39] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.
- [40] T. S. Jaakkola, "Tutorial on variational approximation methods," in *Advanced mean field methods: theory and practice*. MIT Press, 2000, pp. 139–160.
- [41] M. Wainwright, T. Jaakkola, and A. Willsky, "A new class of upper bounds on the log partition function," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2313–2335, July 2005.
- [42] M. Wainwright and M. Jordan, "Log-determinant relaxation for approximate inference in discrete Markov random fields," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2099–2109, June 2006.
- [43] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *Journal of Machine Learning Research*, vol. 6, pp. 905–936, May 2005.
- [44] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.
- [45] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [46] S. Lauritzen and D. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 50, no. 2, pp. 157–224, 1988.
- [47] D. Anguelov, P. Srinivasan, H.-C. Pang, D. Koller, S. Thrun, and J. Davis, "The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces," *Advances in Neural Information Processing Systems*, vol. 17, pp. 33–40, 2004.
- [48] J. M. Mooij, et al, "libDAI 0.2.4: A free/open source C++ library for Discrete Approximate Inference," <http://www.libdai.org/>, 2010.
- [49] M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali, "Simple deterministic approximation algorithms for counting matchings," in *Proc. 39th annual ACM symposium on theory of computing*. New York, NY: ACM, 2007, pp. 122–127.
- [50] D. Musicki and B. La Scala, "Multi-target tracking in clutter without measurement assignment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 877–896, July 2008.
- [51] B. Bakhtiar, H. Alavi, and F. Amoozegar, "Efficient algorithm for computing data association probabilities for multitarget tracking," in *Automated Target Tracking*, F. A. Sadjadi, Ed. Orlando, FL: SPIE, 1996, vol. 2756.